Making an application to the UK Medical Education Database (UKMED)

2022
Contents

What is UKMED 03
Prior to making an application 05
Making an application 07
Glossary 09
FAQs 10
Appendix 1 - Example of a good application 12
What is UKMED

The UK Medical Education Database, or UKMED, is a resource for researchers to better grasp how medical students, trainees and doctors move through medical education and their future careers. By linking data from pre-university qualifications to course performance and postgraduate training, along with various other measures, it is possible to understand the factors that affect how doctors progress through different training pathways.

Access to the database is controlled through an application process. Applications are taken twice yearly, with information about future application rounds released through the UKMED website and newsletter alerts.

UKMED is run jointly by the Medical Schools Council and the General Medical Council (GMC). Its pilot phase ran from 2015 to 2016 and following successful evaluation from stakeholders it has been operating since 2017.

Stakeholders

- Academy of Royal Colleges
- Biomedical Admissions Test
- British Medical Association
- Conference of Postgraduate Medical Deans of the United Kingdom
- General Medical Council
- Graduate Medical Schools Admissions Test
- Higher Education Statistics Agency
- Medical Schools Council
- NHS Education for Scotland
- NHS Health Education England
- Northern Ireland Medical & Dental Training Agency
- UK Foundation Programme Office
- University Clinical Aptitude Test
- Universities and Colleges Admissions Service
- Wales Deanery
Who can apply for a dataset from UKMED?

The UKMED application process is open to all applicants, as long as you can demonstrate appropriate preparation and meet the terms of the UKMED data sharing agreement (a sample data sharing agreement can be found in this document). UKMED data are for research purposes only and cannot be used to make decisions about individual data subjects.

You do not have to be a senior member of staff to apply. For example, medical students and trainees with experience of quantitative data analysis are welcome to apply, as long as you have a named supervisor from an academic institution or relevant organisation and can commit the time. The supervisor will co-sign the contract for accessing UKMED data for research. Accountability lies with the supervisor and their associated organisation for the supervision of the student’s research activities, including compliance with the clauses set out in the contract.
Prior to making an application

The research team

1. You should ensure that within your research group, you have identified a Lead Researcher, who is responsible for report submission to the UKMED Research subgroup and leads the project.

2. Within your group, it is crucial to have a statistical expert.

3. Depending on your research question(s) and complexity of the project, you may wish to approach the relevant data provider to request assistance in interpreting the data. We may be able to provide a contact for you.

4. Researchers who wish to undertake a research project where students are part of the team and plan to use the UKMED data for work required by the course, such as a thesis or dissertation are advised to make an application to UKMED prior to when the student will need to complete the work for their course so that we can ensure the data are available when required.

Stage 1 - before making an application

1. All prospective research applicants new to UKMED should complete an initial draft application form containing details about their research question, methods and analysis. You should send this draft to the UKMED team at least one month before the application window opens. You can contact the UKMED team at info@ukmed.ac.uk

2. You will then be able to receive feedback on this draft application form. It is highly recommended you complete this step to strengthen your application and understand which areas need improvement ahead of the live application window.

3. When reviewing the draft, we will report on whether there are sufficient cases with the requested data and when more data will become available.

Things to note

• It may be useful to review GMC Entrant Profiles to help inform your research question. The draft application form also allows opportunities for feedback on your question, eg an estimate of numbers available.

• You should review the scoring criteria before submitting your draft application to ensure you meet requirements (eg governance).

• Data requested need to be specific to the research questions that are being asked – data not relevant to the question will not be released.
• It is worth considering inclusion of demographic data available in UKMED where this is relevant.

• It is recommended you look at previous accepted applications to check how similar research topics are being addressed. UKMED does not reject applications based on the topic having already been investigated.

• Prepare a realistic timeline. From applying to completing the project usually takes at least 12 months.

• UKMED does not fund research projects. As part of your application, you will need to demonstrate how you plan to resource your research, for example demonstrate that your team have sufficient time within their job plans. Some of the organisations you could approach to explore funding are NIHR, the relevant Royal College or professional body.

Using your own datasets

You may be able to use your own datasets, but you will need to submit these data for matching within UKMED. In order to do this, your data will need to have suitable identifiers (eg GMC numbers, DOB) and you will need to have an appropriate privacy notice on your data collection materials. More information can be sought in this document.

Ethics approval

Researchers are advised to contact their local ethics committee to clarify whether they would need to complete a review process for a project using UKMED data. UKMED can provide further assistance on request.
Making an application

Stage 2 – Making an application

1. All applications are submitted through the UKMED website (www.ukmed.ac.uk). Only one researcher (usually the Lead researcher) should make an account and submit an application. Applications can only be submitted during the application window.

2. The closing date/s for submitting completed applications will be published in advance on the UKMED website.

Stage 3 – Review process

1. Following submission, the UKMED Research Subgroup Members score completed applications independent from each other against the scoring criteria. Members will be able to abstain from scoring a domain if they register conflict of interest or do not have the expertise pertinent to the specific case.

2. Upon completing independent scoring, members of the subgroup will meet and make recommendations to the UKMED Advisory Board. This includes recommendations for approved applications, approved with minor/ major changes and rejected applications.

3. If your application is rejected, but invited for a re-submission, you are able to re-submit within one month of notification. The Chair of the Research Subgroup will re-review your application to see if criteria have been met.

4. The dates for each process are published on the UKMED website once the application window is opened. The lead time between making an application, reviewing and approving applications, confirming the specifications for the data extract, and release of the data can be long. This whole process can take up to three months.

5. In case of approved applications, the title, summary and contact details of the lead researcher as recorded on the application form will be published on the UKMED website.

6. Under most circumstances applications that are approved at the same date will receive their data extracts at the same time. In the event of there being more requests than it is possible to resource, the GMC will work with the UKMED contributors with updates to the UKMED Advisory Board to develop a prioritisation approach.

Stage 4 – Finalising the data specification

1. Once approved, the research team will work with the GMC’s UKMED Data analyst to complete a final specification of the dataset to be used in the research. This
specification will be included in the data sharing agreement.

Stage 5 – Data sharing Agreement

1. Once the specification is finalised the GMC as Data Controller will issue a Data Sharing Agreement. This will contractually restrict the research team’s use of the data to that required for the completion of the research outlined in their approved proposal.

2. It is important to note that the data cannot be used to support measures or decisions with respect to particular individuals and cannot be processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject.

3. The GMC require one signed Data Sharing Agreement per organisation the research team is associated with. This can be signed by the lead researcher, but some organisations prefer that is signed by a Head of department or similar person.

Stage 6 – Accessing data and de-identification

1. The GMC will prepare an extract of data that has been tailored for the research question you are asking. This data will be pseudonymised and moved to a secure environment called the ‘Safe Haven.’ Researchers will conduct their work in the Safe Haven.

2. Your Safe Haven access includes major statistical packages. To complete your work, please make sure you check that the software you plan to use is available within the Safe Haven or request for it to be added with sufficient notice.

3. On completion of the project the GMC as Data Controller will take responsibility for the secure archiving of relevant analysis files for a period of five years, including a copy of the extract provided.
Glossary

Data Dictionary
The document listing all variables available in UKMED grouped into tables.

Safe Haven
The Safe Haven environment is a web accessible Virtual Desktop Environment (VDI) that allows secure remote access to research data provided by HIC Services. It is based on the VMWare View Horizon VDI technology.

UKMED Research Group
A group of experts providing academic support to the UKMED Advisory Board.

UKMED Advisory Board
A group composed of UKMED data contributors.
FAQs

1. What data are held in the database?

For the current data, please consult the data dictionary. The data are comprised of:

Data from the Higher Education Statistics Agency – HESA

- The medical schools attended, including whether the student successfully completed an undergraduate medical course
- Demographic data including ethnicity and measures of socio-economic status
- Data on the students’ entry qualifications (A-levels/Highers and equivalents)

Data from aptitude test providers used by medical schools in their selection processes

- University Clinical Aptitude Test (UCAT)
- Graduate Medical School Admissions Test (GAMSAT)
- The BioMedical Admissions Test (BMAT)

Data from the University and College Admissions Service (UCAS)

Data from applications to foundation training

- Situational Judgement Test scores
- Educational Performance Measures

Data on postgraduate markers of trainees’ progression

- Annual Review of Competence Progression (ARCP) outcomes from deaneries and Local Education Training Boards
- Outcome of applications to nationally recruited specialty training programmes
- Royal College Membership exams

Fitness to practise data (held by GMC)

- Declarations of fitness to practise made on application for provisional registration with a licence to practise
- Fitness to practise data

Data on doctors’ experiences of training from the National Training Survey (held by GMC)
For more information, see the full list of existing datasets. UKMED is currently inviting recommendations for new research questions and associated datasets. Detail on the latest datasets to be added can be found at Potential new datasets.

2. How are data collected?
Data are collated from pre-existing databases used to administer entry to medical school, entry to postgraduate training programmes, the management of postgraduate training programmes and the evaluation of postgraduate training programmes.

3. Who is responsible for UKMED’s compliance with GDPR?
UKMED is a partnership between various organisations. The governance structure of UKMED means the GMC is the sole data controller under the UK GDPR. The GMC is responsible for ensuring compliance with the data protection legislation and responding to requests about the handling of personal data.

4. Can I use the data outside of the Safe Haven?
No, the data will only be available within the safe haven.

5. How is UKMED data updated?
Data are received annually from UCAS and HESA. The GMC’s postgraduate collections are annual. All collections are retrospective, the GMC receives the data after the UCAS cycle, academic year or training year has finished.

Useful links
- Application form for draft completion
- 2022 submission and review dates
- Governance
- Data Dictionary
- UKMED Research Process
- Accepted applications
- Linking your own data to UKMED
Appendix 1 - Example of a good UKMED application

Title
Performance on Situation Judgment Tests and Risk Fitness of Practise Issues in UK Medical Students

Summary
The use of situational judgement tests to evaluate 'non-academic attributes' has rapidly been rolled out in medical selection. There is overall meta-analytic evidence that the scores from such tools are generally valid and add value in selection decisions above measures of knowledge or cognitive ability (Webster, Paton et al. 2020). Nevertheless, this latter evidence synthesis highlighted that the validity coefficients from SJT evaluations based on postgraduate settings were statistically significantly larger than those reported for undergraduate settings (β = 0.21, p<0.001). Previously we have demonstrated a statistically significant association between performance on the SJT used in the allocation process for the foundation years and the risk of reporting a conduct related fitness to practice issue at provisional registration with the General Medical Council (Paton, Tiffin et al. 2018). However, these SJTs were administered after medical student selection into undergraduate studies had already taken place. Moreover, as a result of this initial study, changes in the way that fitness to practice declarations were made have been enacted.

These are now intended to be validated by medical schools, rather than purely relying on the self-report of the student at provisional registration. The extent to which the self and medical school reports agree will be evaluated shortly by the GMC.

The situational judgement test currently used as part of the University Clinical Aptitude Test (UCAT) was first piloted in 2013 and used within selection in 2014. Validation studies highlighted relatively modest, though statistically significant, correlations between the scores on this SJT and subsequent tutor ratings (Patterson, Cousans et al. 2017). An independent analysis also highlighted that the six different forms of the test may have been working slightly differently in the pilot version (Tiffin and Carter 2015).

Thus, it appears timely to conduct a study specifically in relation to SJT performance and the risk of undergraduate conduct problems. That is, if SJTs are to be effective in selection then they should have a positive footprint in reduced rates of professionalism lapses in both medical students and doctors. Moreover, it is established that lapses of professionalism in undergraduate studies are a risk factor for subsequent censure for fitness to practice issues in qualified doctors (Papadakis, Teherani et al. 2004). If we report positive findings then it would be clear that the UCAT SJT is serving an important role in helping to ensure public protection from unprofessional behaviours in medical graduates and doctors, which are generally linked to personal qualities, rather than lack of clinical knowledge or skills (Tiffin, Paton et al. 2017).

This study is how feasible due to changes in the way that fitness to practice lapses have now been recorded over the last three years, and the maturity of UKMED; there are now data in UKMED for medical students who both completed the UCAT SJT in 2013, 2014 and 2015 and have completed a provisional registration declaration; by October 2020 there are estimated to be around 7000 entrants with SJT scores and FTP declarations present. Therefore, we are now in a position to meaningfully model the potential impact of introducing the SJT into undergraduate medical selection. In particular it would be useful to understand where the optimum, and practical, threshold for performance on the UCAT SJT might lie. Thus, our findings would have clear implications for the way that the SJT is implemented within undergraduate medical selection, both within the UK and Australasia.

References

Research questions
1. Is there a relationship between performance on the UCAT SJT and the risk of declaring a conduct related fitness to practice issue during the undergraduate years?
2. If such a relationship exists, to what extent is this independent of other key selection measures; namely performance on the cognitive components of the UCAT, and school academic performance at advanced level? That is, do the SJT scores demonstrate incremental predictive value in this context?
3. If such a relationship exists what might be the impact on the number of fitness to practice declarations made of taking varying cut-offs for SJT performance, as a barrier to medical school entry?
4. We will also assess evidence for convergent validity of the UCAT SJT by evaluating the degree of correlation with the Foundation Programme allocation SJT scores.
Data required

<table>
<thead>
<tr>
<th>Data Item</th>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCAT_MOTHERTONGUE</td>
<td>N/A</td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>UCAT_LANGUAGE_FLUENCY</td>
<td>N/A</td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>SCHOOL_STATUS</td>
<td></td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>NATIONALITY_SRC</td>
<td></td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>PMQ_WORLD_REGION_AT_R ECU</td>
<td>N/A</td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>NSSEC</td>
<td>N/A</td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>SEC</td>
<td>N/A</td>
<td>National Statistics socio-economic five-point scale classification (1 = managerial and professional occupations, 2 = intermediate occupations, 3 = small employers and own account workers, 4 = lower supervisory and technical occupations, 5 = semi-routine and routine occupations)</td>
</tr>
<tr>
<td>POLAR3_SRC</td>
<td>N/A</td>
<td>Source system for column POLAR3</td>
</tr>
<tr>
<td>POLAR3</td>
<td>N/A</td>
<td>Low-participation neighbourhoods (POLAR3) - Classification of areas for young participation rates in higher education. Participation rate in higher education associated with the student’s domicile postcode.</td>
</tr>
<tr>
<td>POLAR2_SRC</td>
<td>N/A</td>
<td>Source system for column POLAR2</td>
</tr>
<tr>
<td>POLAR2</td>
<td>N/A</td>
<td>Low-participation neighbourhoods (POLAR2) - Classification of areas for young participation rates in higher education. Participation rate in higher education associated with the student’s domicile postcode.</td>
</tr>
<tr>
<td>SCHOOL_TYPE_SRC</td>
<td>N/A</td>
<td>Source system for column SCHOOL_TYPE</td>
</tr>
<tr>
<td>SCHOOL_TYPE</td>
<td></td>
<td>Describes which type of school the doctor mainly attended between the ages of 11 and 16. MESA State school marker</td>
</tr>
<tr>
<td>BME_INT</td>
<td>N/A</td>
<td>Higher level ethnicity coded as an integer.</td>
</tr>
<tr>
<td>BME</td>
<td></td>
<td>Higher level ethnicity coding: BME or white.</td>
</tr>
<tr>
<td>ETHNICITY_SRC</td>
<td></td>
<td>Source system of the ETHNICITY_L1 and ETHNICITY_L2 data stored for this record.</td>
</tr>
<tr>
<td>ETHNICITY_L1_INT</td>
<td>N/A</td>
<td>Ethnicity Level 2 information coded as an integer.</td>
</tr>
<tr>
<td>ETHNICITY_L2</td>
<td></td>
<td>Ethnicity Level 2 information.</td>
</tr>
<tr>
<td>ETHNICITY_L1</td>
<td></td>
<td>Ethnicity Level 1 information coded as an integer.</td>
</tr>
<tr>
<td>ETHNICITY_L1_INT</td>
<td></td>
<td>Ethnicity Level 1 information coded as an integer.</td>
</tr>
<tr>
<td>HESA_TARIFF</td>
<td>NUMBER(S, Z)</td>
<td>Tariff. - Tariff points based on the qualifications on entry of the student. Tariff data was collected directly from UCAS prior to 2007/08 and was calculated by HESA for those entering from 2007/08 onwards. As such, tariff data will not be directly comparable across this time period.</td>
</tr>
<tr>
<td>GENDER_INT</td>
<td>N/A</td>
<td>Gender coded as integer: 1 = Female and 0 = Male.</td>
</tr>
<tr>
<td>GENDER</td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>BIRTH_YEAR</td>
<td>N/A</td>
<td>Birth year extracted from Date of Birth held on the GMC register. Date of birth is too identifiable for inclusion in extracts.</td>
</tr>
<tr>
<td>BIRTH_MONTH</td>
<td>N/A</td>
<td>Birth month extracted from Date of Birth held on the GMC register. Date of birth is too identifiable for inclusion in extracts.</td>
</tr>
</tbody>
</table>

Methodology

As with our previous study (UKMED P001) the fitness to practice declarations will be categorised into health-related conduct related, according to the descriptions given of the events. Although our primary focus is on the ability to predict conduct related fitness to practise events, we shall also evaluate for a potential relationship with health-related declarations. This is because there is interest in the resilience of medical students and doctors, and better interpersonal skills may be a protective factor in this context. Therefore we will carry out an analysis evaluating the univariable relationship between the UCAT SJT scores and the odds of a fitness to practise declaration. A series of multivariable models will also be tested. These will include cognitive ability, as estimated by the most recent UCAT total score recorded for the student in UKMED, and secondary school performance, as assessed by the UCAS tariff score at university application. If a statistically significant univariable relationship exists, then a series of simulations will be conducted (see Tiffin et al. 2016, for a description of this approach https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-016-0682-7). These simulations will aim to estimate the ‘number needed to reject’ for one or more putative cut-offs for the SJT performance. This simulation will involve describing how the number of conduct related FIP events would change in the sample for different SJT cut-off scores/bands. This should provide an estimate of the likely real-world impact of changing the SJT admission criteria, albeit a crude one. Sensitivity analysis to evaluate the missing at random (MAR) assumption for missing predictor variables will be conducted via multiple imputation.

Commented [FY9]: This is reviewed to ensure you have sought the appropriate data to answer the question. If helpful consult with Daniel Smith in advance of your submission. But not last minute please!

Commented [FY10]: Detail is critical here so we can be confident that you are aware of the complexities and opportunities of using the data well. Sometimes we ask that your team is expanded to include someone with relevant insights when this appears lacking - e.g. a particular training pathway or assessment. Consider if this might help your team?

Specifically, please consider sample size adequacy, handling missing data, any issues you see or judgments you will need to make such as categorisations and outline your intended statistical approach. Ensure Privacy Notices are addressed if importing data and if relevant, check you have considered correction for prior attainment and use of precalculated fields.
Analysis proposed

Data would be linked, checked and cleaned. The univariable potential predictors (used at undergraduate selection; SJT, UCAT cognitive scores and school academic performance) of a positive report on a Fitness to Practise Declaration at provisional registration would be explored. FFP declaration categories will be collapsed according to substantive considerations and the numbers of individuals falling into each group. Binary logistic regression will be conducted to evaluate the relationship between admission characteristics and FFP declaration status. In some cases multiple declarations will be made by single individuals. Thus it may necessary to adjust for such dependency using a multilevel modelling or generalized estimating equation (GEE) framework with logit link functions. In this case a stepwise approach to multivariable model building will not be used. Rather, in order to evaluate the potential incremental validity of the UCAT SJT scores these will be entered alongside secondary school academic performance and UCAT cognitive subtests course. In order to compare the qualities of the UCAT SJT with the SJT used in the foundation programme application process the analyses will be repeated, except substituting former for the latter. Sensitivity analysis to evaluate the missing at random (MAR) assumption for missing predictor variables will be conducted via multiple imputation. This should only apply to missing data in relation to education attainment.

Sensitivity analyses also be conducted to evaluate the extent that including different forms of the UCAT SJT are likely to impact on the results. Simulations will be carried out for the cohort by taking possible cut-offs (to be decided by the earlier analyses) for the UCAT SJT scores and evaluating the extent to which the impact on the number of conduct related fitness to practice declarations in each situation. In this way we will be able to estimate the ‘number needed to reject’ for the UCAT SJT at a particular cut-off. The number needed to reject represents the ratio of individuals without an undesirable outcome needed to be rejected to avoid admitting one candidate likely to have this negative outcome. In this case, unlike previous similar work estimating this index will be performed on the admitted cohort, rather than simulated with medical school applicants. This is because many of the applicants may have other reasons why they would not be likely to be admitted to medical school, aside from the SJT score, such as a relatively low score on the cognitive subtests of the UCAT. Moreover the UCAT SJT was not used in selection 2013 and was largely only used very cautiously in 2014 and 2015. Nevertheless, there may have been some indirect restrictions range, in that those who perform interviews were also likely to have performed well on the UCAT SJT. This is a potential limitation.

Timeline

Proposed Date: 2020-11-25

Duration: 6 month(s)

1st November 2020 to 31st December 2020: Data received and cleaning and management carried out 1st January 2021 to 1st February 2021: Data analysis conducted 1st February to 31st March 2021: Writing up report

Proposal for dissemination

The study will be written up for publication in an appropriate high-impact peer-reviewed journal. The results will also be separately reported to key stakeholders. These will include the Medical School Council, Work Psychology Group (who designed the UCAT SJT for delivery by Pearson Vue, and are aware of this proposed study). The lead author is a member of the UCAT research and development group and the results will also be presented to the UCAT Board and Pearson Vue. Any incidental, relevant findings in relation to the foundation programme SJT will also be presented, as with similar research, to the UKFPO for consideration.
### Researchers and partners

| Name                  | Org          | Role                | Proposed Role | Experience with managing, cleaning and organising large datasets (>5000 cases) | Experience with statistics and data modelling | What software do you have experience and expertise in using? | CV                  | Commented [FY14]: We are seeking assurance that the study is feasible: 
- Subject matter expertise
- Analytical expertise
- Capacity/time
Please address all, or else we are likely to seek clarification. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctor Paul Tiffin</td>
<td>University of York</td>
<td>Lead Researcher</td>
<td>Lead and project supervision. Will lead on data cleaning, management and supervision and the help perform the analyses. Lead on writing up the results for publication.</td>
<td>Extensive experience of managing, cleaning and analysing large datasets using conventional statistical techniques and machine learning approaches. I have extensive experience of conventional statistical approaches as well as machine learning. This includes the use of multilevel models, structural equation modelling, kernel variable selection and the application of time series analysis.</td>
<td></td>
<td>RStata, R, Python</td>
<td>Paul is a Reader in Psychiatric Epidemiology/Consultant Psychiatrist and quantitative methodologist. He is a member of the UKMID Research Subgroup and led UKMID PWI: a study into PV in undergraduates. Paul has published over 80 journal papers, mainly as first or named author and raised over £2.5M in research funding as a lead and co-applicant.</td>
</tr>
</tbody>
</table>
Doctor Lewis Paton  
University of York  
Researcher Dr Paton will provide additional statistical assistance to the data management, cleaning and supervisory input to analysis and write up.

Extensive experience with UKMED, UCAT, CPRD and other large datasets.

MPhil in Bayesian Statistics (Durham University), BSc in Mathematics (Class I, Durham University)

Extensive postdoctoral experience in data modelling and statistics, including machine learning approaches.

Included in communication: Yes  
Access to safe haven: Yes

Mr Daniel Smith  
CMC  
Researcher Mr Smith will advise on data requirements for the study and also perform pre-study evaluation of the degree to which FTP declarations coincide with FTP issues reported by the medical schools. He will also categorise FTP issues if required.

Mr Smith is the data development manager in the main CMC UKMED analyses. Experience of analysing large datasets and has a postgraduate diploma in Health Informatics from the University of Sheffield.

SPSS, Excel R

Included in communication: Yes  
Access to safe haven: Yes

Doctor Emily Sanger  
University of York  
Researcher Dr Sanger will perform data analysis under supervision and prepare the final draft of the UKMED report and paper for publication.

Will gain experience under supervision with UCAT, UKMED.

Postgraduate Certificate in Health Research and Statistics, University of York, 2018 - Present

SPSS, Excel Gpower

Included in communication: Yes  
Access to safe haven: Yes

**Funding**

This study will be conducted as part of PAT’s (lead author’s) NIHR career development fellowship programme of work on medical selection. This ends officially December 31, 2020. However, data analysis and writing up will be assisted by Dr Emily Sanger, academic foundation programme doctor, who will be involved with this project almost full-time during December 2020 to April 2021. In addition, Lewis Paton is funded one day a week by the UCAT Board to perform research relevant to UCAT’s aims and objectives.

Commented [FY15]: Funding is not a requirement but when cited we will seek confirmation, so best to provide in advance where able. We are seeking to ensure the work will be feasible according to the proposed timeline.